

# Usage reporting on recorded lectures using educational data mining

Citation for published version (APA):

Gorissen, P., Van Bruggen, J., & Jochems, W. (2012). Usage reporting on recorded lectures using educational data mining. *International Journal of Learning Technology*, 7(1), 23-40.  
<https://doi.org/10.1504/IJLT.2012.046864>

**DOI:**

[10.1504/IJLT.2012.046864](https://doi.org/10.1504/IJLT.2012.046864)

**Document status and date:**

Published: 01/01/2012

**Document Version:**

Early version, also known as pre-print

**Please check the document version of this publication:**

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

**General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

<https://www.ou.nl/taverne-agreement>

**Take down policy**

If you believe that this document breaches copyright please contact us at:

[pure-support@ou.nl](mailto:pure-support@ou.nl)

providing details and we will investigate your claim.

Downloaded from <https://research.ou.nl/> on date: 05 May. 2023

**Open Universiteit**  
[www.ou.nl](http://www.ou.nl)



---

## Usage Reporting on Recorded Lectures

---

Pierre Gorissen\*

Fontys University of Applied Sciences, P.O. Box 347,  
5600AH, Eindhoven, the Netherlands

E-mail: P.Gorissen@fontys.nl

\*Corresponding author

Jan van Bruggen

Fontys University of Applied Sciences, P.O. Box 347,  
5600AH, Eindhoven, the Netherlands

Open University of the Netherlands, P.O. Box 2960,  
6401DL, Heerlen, the Netherlands

E-mail: J.vanBruggen@fontys.nl

Wim Jochems

Eindhoven University of Technology, P.O. Box 513,  
5600MB, Eindhoven, The Netherlands

E-mail: w.m.g.jochems@tue.nl

### Abstract:

Most research of the use of recorded lectures by students is based on surveys. In this article we will report on the further analysis of student's use of recorded based on methodological triangulation. We will describe the process of data pre-processing where a number of data sources are combined after which the data is cleaned, removing outliers and data not relevant to the research.

The analysis looks both at the data from a high level point of view and on the more detailed level for a single course. We will show differences and similarities in the responses by the students about their use of the recorded lectures and the data logged by the systems that serve the recorded lectures to the students.

**Keywords:** recorded lectures, weblectures, educational data mining, learning analytics, video, lectures, lecture capturing, reporting, learning technology.

**Biographical notes:** Pierre Gorissen is senior consultant at the educational development and research department of the Fontys University of Applied Sciences. He is responsible for projects related to the use of ICT in education. He is a PhD candidate at the Eindhoven School of Education, part of the Eindhoven University of Technology. His research focuses on the use of recorded lectures by students, their goals for use and ways to support that use.

Jan van Bruggen has a PhD in educational technology. He is associate Professor at Fontys University for Applied Science and at the Open University of the Netherlands (OUNL). At Fontys he heads a lectorate on research and development in educational applications of ICT, such as student support systems. At OUNL he leads one of the themes in the Learning Networks program.

Wim Jochems received his Master degree in Educational Psychology and Methodology from Utrecht University, and his PhD in Technical Sciences from Delft University of Technology (TU Delft). In 1998 he became Dean of the Educational Technology Expertise Centre at the Open University of the Netherlands (OUNL) and full professor in Educational Technology. Currently, he is Dean of the Eindhoven School of Education at Eindhoven University of Technology.

## **1 Introduction**

The lecture has been around for hundreds of years. The lecture method is the most common form of teaching in institutions of higher education throughout the world (Behr, 1988). From a technology point of view, a lecture can be seen as a multimedia activity where the lecturer uses a combination of speech, writing and projected video and images to convey a message to the students (Abowd et al., 1996). Recorded lectures use computing technology to facilitate the automatic capture and integration of and access to the media used during a lecture (Abowd et al., 1998a, Brotherton and Abowd, 2004, Abowd et al., 1998b). These are integral recordings of lectures providing students with the opportunity to review lectures online at their own pace and at their preferred time and place. An increasing number of universities choose to support the learning of their students by providing online recordings of lectures (Leoni and Lichti, 2009).

Most of the research in this area is based on surveys and verbal reports by students of their use of recorded lectures. Traphagan (2005, 2006), Veeramani and Bradly (2008) and Gosper et al. (2008) report that most students prefer courses accompanied by online recordings of the lectures. Gorissen, van Bruggen and Jochems (submitted) conducted a survey into the use of recorded lectures by students at two universities in the Netherlands. Students report using recorded lectures as an effective and versatile study tool. They use recorded lectures as an incidental or structural replacement for missed lectures as well as an aid for specific purposes, such as exam preparation. The research suggests that students use the recordings differently at specific moments and for specific purposes. Sometimes they watch the entire recording; at other times, they only watch parts of it. A large number of students, however, report that they view most of the recorded lectures. The study shows that there is a need for further analysis of the student's use of recorded lectures by comparing their responses to other data sources. Methodological triangulation (Denzin, 2006) can help to increase the reliability and validity of these results. In this article we report on a study into usage reporting as a method to improve our understanding of the use of recorded lectures by students at a university in the Netherlands.

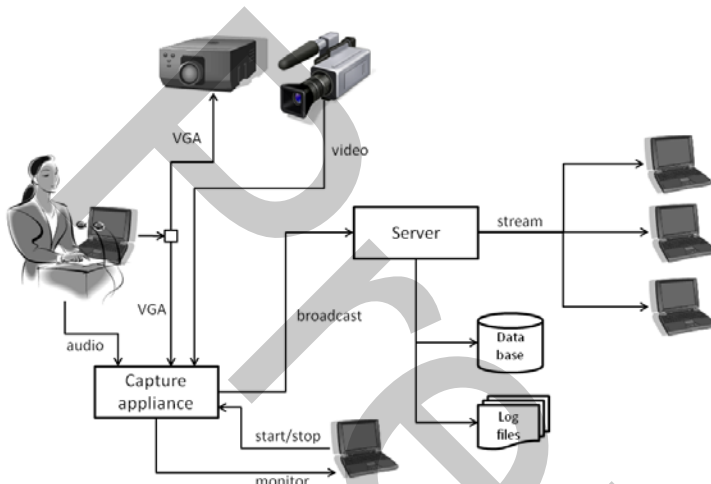
The systems that provide accesses to recorded lectures record a significant amount and variety of data on what students are doing within the system. In this study we use a process of educational data mining to further analyse the use of recorded lectures by students. Data mining is the process of discovering patterns in data. This process can be automatic or (more usually) semiautomatic. The data must be present in substantial quantities (Witten and Frank, 2002). The patterns sought for must be meaningful in that they lead to some advantage, usually an economic advantage. In an educational setting, patterns are sought that help to better understand students and the settings in which they learn (Baker et al., 2010).

The purpose of our analysis is to study differences and similarities between the use of recorded lectures as reported by students, and the actual use, as determined by analysing the log data.

## 2 Lecture capturing

There are many different system architectures in use for broadcasting and recording lectures. Figure 1 shows an example of a lecture capturing system (LCS) providing both live and on-demand access to recordings.

**Figure 1** A lecture capturing system (LCS) architecture



The figure shows a common setup for a LCS, with a dedicated capture appliance in the lecture hall, capturing audio, video and the VGA signal of the lecturer's laptop or desktop. The capture appliance can be started, stopped and monitored remotely by a system administrator. The capture appliance automatically synchronises the recorded audio, the video and the VGA signal into a single interface (Figure 2) stored on the server and can be viewed live or on demand by the user using a web browser. Information about the recorded lecture, such as the title, the name of the lecturer, the name of the course and the date and time of the recording, also is stored on the server in the database. Access to recorded lectures is registered in log files (text files) on the server.

**Figure 2** Recorded Lecture User Interface example



### 3 Method

We analysed recordings created by the Eindhoven University of Technology (TU/e) in the Netherlands. The university uses the Mediasite LCS by SonicFoundry (Sonicfoundry). They have five capture appliances they use to create recorded lectures. Currently the TU/e Mediasite repository holds about 8,000 recordings with a total length of over 5,000 hours of video. All recordings are available online; students can view them in their browser, both at the university and from home. No downloadable versions of the recordings are provided. The recorded lectures are typically about 40-45 minutes long, containing one full lecture each. The number of recordings per course depends on the number of lectures for that course; the average is 16 recordings for a single course, and the maximum number of recorded lectures for a single course is 54.

Before we can analyse the data, we need to perform a number of steps to prepare the dataset for analysis. This process is called ‘data pre-processing’ (Sheard, 2011) and consists of the following steps: combining data sources, determining missing entries, removing irrelevant data, user identification, session identification and removing outliers.

The first part of the analysis will be conducted on the resulting filtered total dataset. The second part of the analysis focuses on one single course. This course is part of the set of courses surveyed in the study presented by Gorissen, van Bruggen and Jochems (submitted). It is the course with the most views for recorded lectures and thus with the biggest individual data set.

**Figure 3** Data pre-processing steps

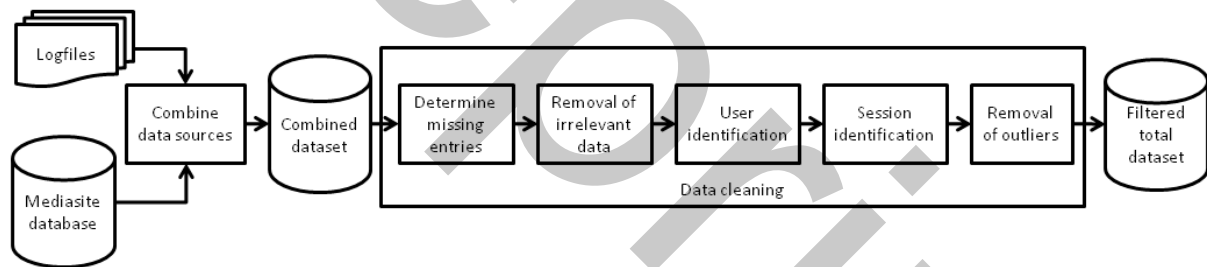


Figure 3 shows the data pre-processing steps. First, the two data sources are combined, and then the data is cleaned, providing a filtered total dataset.

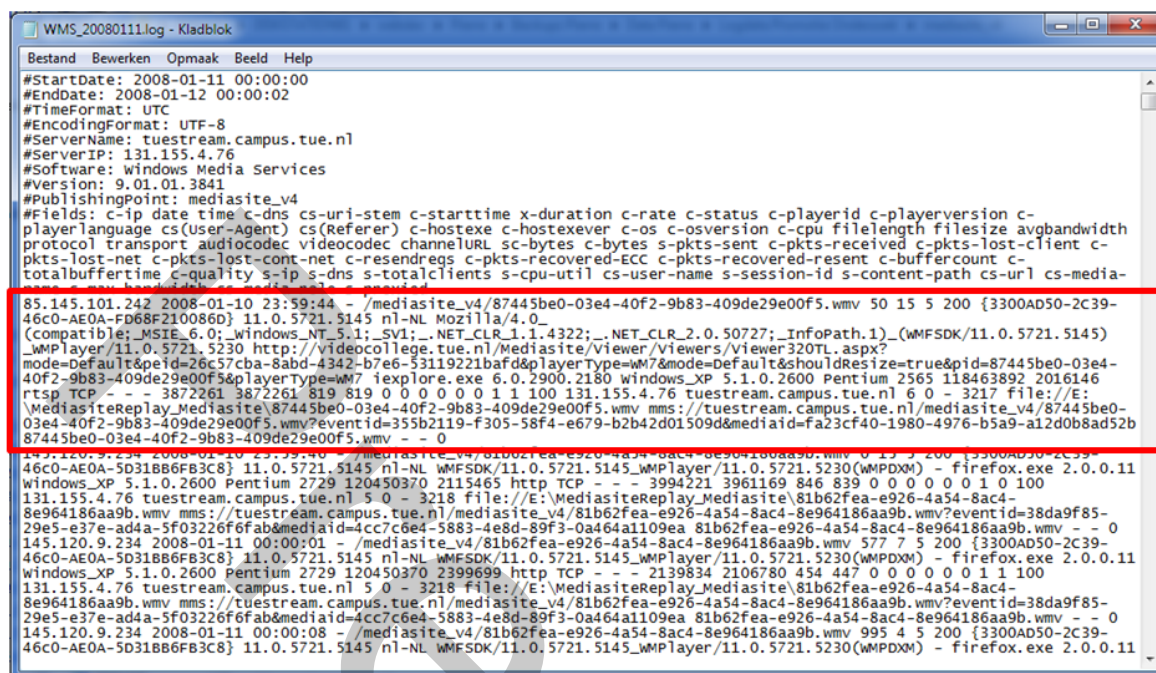
#### 3.1 Combine data sources

For the analysis we use a combination of two data sources: one consists of the database of the Mediasite LCS used to record the lectures. This Mediasite database is stored in a Microsoft SQL Server and contains a total of 92 tables with information about the lectures that have been recorded (title, course, recording data /time), the lecturers/presenters, students' user ids, staff and others that viewed the recordings, and other data that Mediasite LCS needs. A copy of this database was created to use during the analysis.

The second data source consists of the log files of the video server (Microsoft Windows Media Service) that streams the video of the recordings to viewers. The log files consist of a set of text files created by the system, one single file per day. Each file contains a number of log entries, one for each request to the server. A request is an uninterrupted stream of video sent to the viewer. Whenever a viewer skips ahead or back in the video or jumps to a new recording, a new entry (line) in the log file

is added. Figure 4 shows an example of such a log file with a single log entry marked. The total set of log files contains more than 1.5 million log entries.

**Figure 4** Microsoft Windows Media Service log file



The format of the log file used by Microsoft Windows Media Service (WMS) has been documented by Koyun (2007). We use only a portion of the available data (see Table 1).

**Table 1** Part of the data available in each Microsoft Windows Media Service Log file row.

| Element     | Description  | Example  |
|-------------|--|--|
| c-ip        | The source Internet Protocol (IP) address of the connected presentation viewer.                                      | 85.145.101.241   |
| date        | Date of the log entry. This is the date the request described in the entry started.                                  | 2011-01-10   |
| time        | Time of the log entry. This is the time the request described in the entry started.                                  | 23:59:44   |
| cs-uri-stem | The path (requested URL without the schema, host, port number, and question mark) to the content that was requested. | /mediasite_v4/74875f80-ac29-4c79-a935-b99ec476d01e.wmv |
| c-starttime | Timestamp (in seconds) indicating the point in the stream when the presentation viewer started to render content.    | 50   |
| x-duration  | Length of time (in seconds) of the data received by the presentation viewer.   | 15   |

The log files were imported into the same SQL Server database as the copy of the Mediasite LCS database. The data was combined so that the c-starttime and x-duration data could be linked to learner

sessions, and IP addresses (c-ip) were linked to user names that had viewed the recorded lectures. In the analysis and reporting, the data were anonymised to ensure the privacy of the users involved.

### 3.2 Data cleaning

The Mediasite database contains data starting March 29, 2004 (the oldest recording still available live online); the log files available for our analysis range from January 9, 2008 up to and including August 31, 2010. For our analysis we needed data from both sources, so we limited the set to the range January 9, 2008 up to and including August 31, 2010.

Further analysis of the dataset showed that the Mediasite database did not contain data for the time period January 19, 2009 – July 27, 2009. This means we were unable to connect the detailed information in the log files to individual users. Therefore, this time frame also was excluded from our dataset.

The TU/e system not only records lectures, but makes recordings of seminars, public speakers and other events that take place during the year as well. We were only interested in the recordings of the lectures, so we filtered out all the other recordings and their visits from the data set.

All users who wish to view recorded lectures are required to log in using their individual user name and password. This means we could identify every unique user that viewed one or more recorded lecture. For our analysis we were interested in the use of the recordings by students. We filtered out all other users from the dataset. This included other staff, the professors/lecturers, and the researchers conducting this analysis.

The aim of the study is to analyse the interaction of the student with the recorded lectures. These interactions are not analysed individually, but in the context of a 'learner session': an uninterrupted period of time during which a learner accesses one or more recorded lectures (Advanced Distributed Learning, 2004). The start of a learner session is determined by the first request that a student makes to view a lecture recording. It is, however, not as easy to pinpoint the exact ending of a session. The Mediasite system does not receive a notification when a student is finished viewing lecture recordings. This can only be determined by the absence of new requests for video.

This problem is similar to the one faced by every website owner and is directly linked to the semi-stateless nature of the World Wide Web. For Web analysis, the session is terminated when an individual has not taken another action on the site within a specified time period (Burby et al., 2007). For the purpose of this analysis, the same method of time-based termination of learner sessions is used. The termination of a learner session is calculated as:

$$\langle \text{learner session end} \rangle = \langle \text{Date of log entry} \rangle + \langle \text{Time of log entry} \rangle + \langle \text{x-duration} \rangle + \langle \text{time-out period} \rangle$$

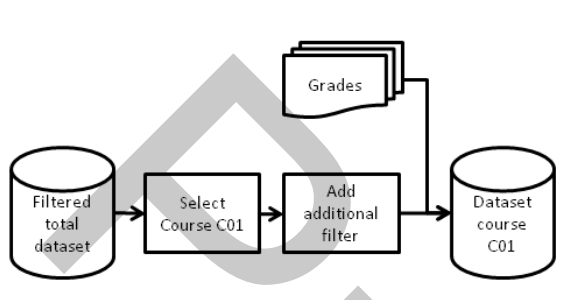
The time-out chosen is usually about 20-30 minutes. The time-out chosen here is 4 hours. We assume that a learner session not only consists of constant viewing of recorded lectures, but that students also review notes, do assignments, read in their textbooks, take a short break within a learner session, etc.

The final step in pre-processing the data is the removal of outliers. We were only interested in learner sessions where the students actually make use of the recorded lectures. Learner sessions shorter than 3 minutes or learner sessions where a total of less than 2 minutes of video has been received by the student are not considered to be actual use of the recorded lectures as part of study activities. These sessions have been removed from the dataset.

### 3.3 Creation of a dataset for detailed analysis

For the detailed analysis of the course with the most learner sessions (course C01), we performed two additional steps (see Figure 5): we filtered out all learner sessions that do not contain one or more of the recorded lectures of the C01 course, and we removed all learner sessions where less than one minute of video was sent for lecture recordings that are part of C01. This filters out learner sessions where course C01 is only a small part of the total learner session.

**Figure 5** Creation of dataset for detailed analysis



For the user analysis for this course, we retrieved all grades for Course C01 and partial grades for C01P01 and C01P02 for the timeframe August 2009 – August 2010 from TU/e’s Student Grade System. The data was imported into the dataset so that it could be used during the detailed analysis of course C01.

## 4 Results for the total dataset

The filtered total dataset contained data on 4,192 lecture recordings, for a total of 263 different courses. It contained 48,539 learner sessions, viewed by 4,927 unique students. The average number of lecture recordings per course is 16, with a maximum of 54 lecture recordings per course. Students watched an average of three different recorded lectures per learner session ( $Mdn = 2$ ,  $SD = 2.6$ ). Table 2 shows the average, median, and maximum number of learner sessions per course, per recorded lecture and per student.

### 4.1 Usage analysis

Use of recorded lectures can be expressed as the amount of video that users received as well as the length of learner sessions. It is not possible to determine whether students actually view the video they receive. They might not be at their desk while the video plays, the video might not be visible (obscured by other windows), or students may not be paying attention. It is, however, certain that they cannot have viewed video that they did not receive. We can determine the maximum amount of video time the student may have viewed, and we know what parts of the recorded lecture students received.

**Table 2** Usage overview per learner session in minutes

|                | Average | Median | Maximum | Minimum |
|----------------|---------|--------|---------|---------|
| Length         | 99      | 50     | 23*     | 3       |
| Received video | 28      | 17     | 10*     | 2       |

\* Hours



Table 2 shows an overview of the differences between the learner session length and the received video per learner session. Note that the minimum amount of both is set by the filter we applied during the pre-processing. The table shows that the maximum length of a learner session is 23 hours. This does not mean that the student was viewing recorded lectures without interruption for 23 hours. It means that the student has been active (navigating or receiving video) with intervals that don't exceed 4 hours. During those learner sessions, the maximum amount of video received in a single learner session was 10 hours. Twenty-nine percent of all learner sessions contain recorded lectures that were created three weeks or less before the time of viewing. Fifty-four percent of all learner sessions use recorded lectures that are two months or older, 23% of all learner sessions use recorded lectures one year or older, and 1% of all learner sessions use recorded lectures that are two year or older.

#### **4.2 User analysis**

We cannot easily compare the individual users at an overall level for all the 263 courses. Most users are only interested in a subset of the recorded lectures. Each week there are new users that use the recorded lectures for the first time, and there are users that stop using the recorded lectures. On average, 18% of all users in a given week use recorded lectures for the first time. Nineteen percent of all students have had one learner session in total, and 38% of all users only viewed recorded lectures in a single week during the two-year time span of our dataset.

By focusing on a single course, we limit our analysis to a more consistent group of students.

### **5 Results for the detailed analysis of course C01**

Course C01 is a course at the faculty of Industrial Engineering & Innovation Sciences of the TU/e. Students that participate in the course come from a number of different departments within the university. Most of the students (66%) are from the Industrial Engineering department, another substantial group of students (23%) is from the Innovation Sciences department. The course consists of an introduction in empirical research. Students learn how to translate real-life questions into research questions, and they learn how to create and evaluate a research design. In the second part of the course, they get hands-on training using SPSS. The course is taught by two lecturers: one for the first part of the course and one for the second part. The first part of the course is taught in Dutch, the second part is taught in English.

Grading for the course is done in two parts, and grades are registered for both individual parts and for the complete course. The first partial grade (C01P01) consists of the combined result for a 90-minute written test and an assignment. The second partial grade (C01P02) consists of a three-hour laptop test. Each partial grade (C01P01 and C01P02) needs to be 5.0 or higher on a 10-point scale. If that is the case, the grade for C01 =  $.5 \times C01P01 + .5 \times C01P02$ . This combined grade needs to be 6.0 or higher on a 10-point scale. Each partial grade can be retried once every year. Partial grades can extend into the next year.

The lecture recordings we examined span the period August 2009 through January 2010. During that period, there were a total of 35 recorded lectures captured as eighteen 2 x 45-minute recordings (the last lecture only consisted of 1 x 45 minutes). During one lecture, Lecture 2a, the audio for the recording failed, resulting in a useless recording 2.5 minutes long.

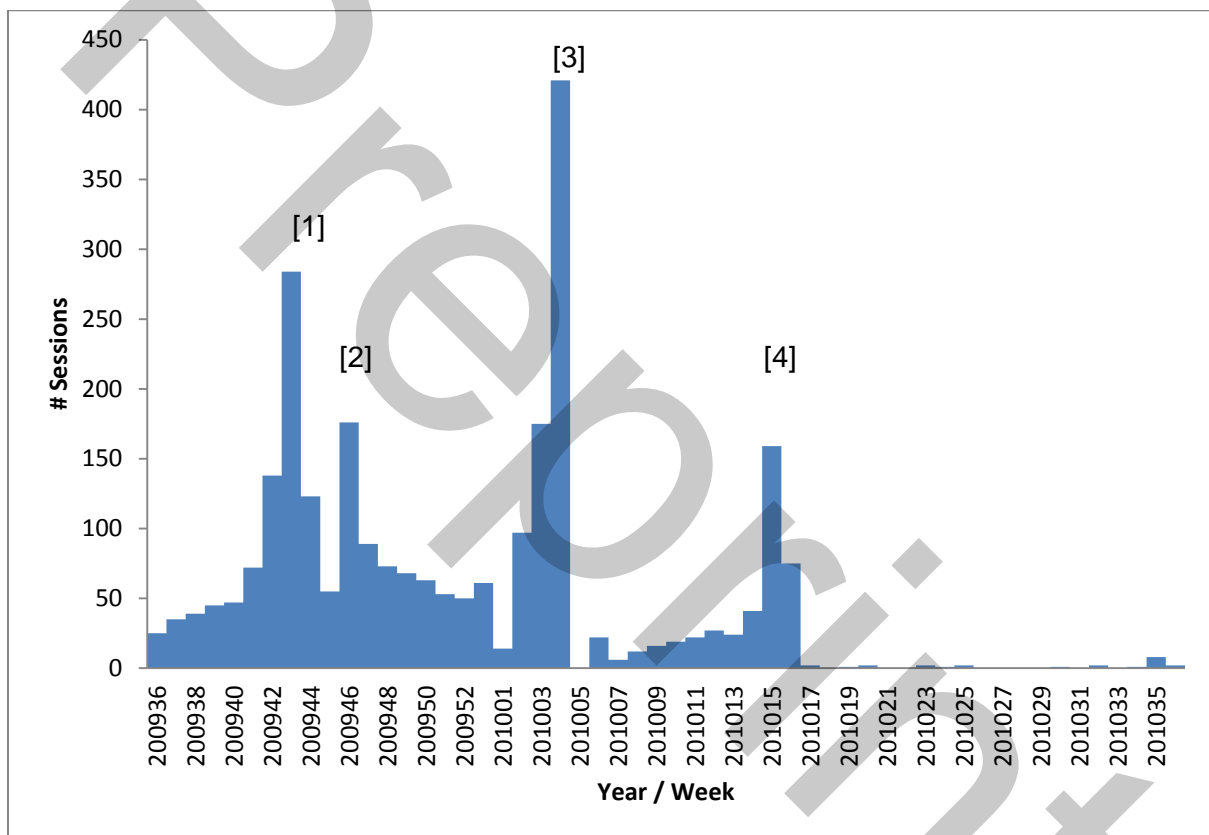
During this detailed analysis of course C01, we will use the created “dataset for the detailed analysis”.

### 5.1 Usage analysis

Figure 6 shows the number of learner sessions per week for the timeframe August 2009 (week 36 of 2009, indicated as 200936) through August 2010 (week 35 of 2010, indicated as 201035).

The graph contains four weeks in which there is a higher-than-average use of the recorded lectures. The first, week 43 of 2009 (200943 – [1]) coincides with the written test for C01P01, the first part of the course. The second, week 46 of 2009 (200946 – [2]), is the week in which the assignment for the first part of the course was due. The third peak in the graph is the largest and occurred during week 4 of 2010 (201004 – [3]). This is the week leading up to the laptop test of the second part (C01P02) of the course. The last peak occurred in week 15 of 2010 (201015 – [4]), which is right before the second opportunity for both the written test for C01P01 and the laptop test for C01P02.

**Figure 6** Number of learner sessions per week



Most of the learner sessions for the recorded lectures of the first part of the course take place within the first two months after they have been recorded (see Table 3).

**Table 3** Unique learner sessions per recorded lecture per month for the first part of the course

| Lecture | Month |    |     |     |    |      |    |    |    |    |    |    |    | Total |
|---------|-------|----|-----|-----|----|------|----|----|----|----|----|----|----|-------|
|         | 2009  |    |     |     |    | 2010 |    |    |    |    |    |    |    |       |
|         | 08    | 09 | 10  | 11  | 12 | 01   | 02 | 03 | 04 | 05 | 06 | 07 | 08 |       |
| 1a      | 7     | 33 | 66  | 2   |    | 4    | 1  | 1  | 1  |    |    | 1  | 8  | 124   |
| 1b      | 4     | 22 | 58  |     |    |      | 1  | 1  | 1  |    |    |    | 4  | 91    |
| 2a      |       | 6  | 5   |     |    |      | 1  |    |    |    |    |    | 2  | 14    |
| 2b      |       | 38 | 127 |     | 1  |      | 1  | 1  | 3  |    |    |    | 3  | 174   |
| 3a      |       | 33 | 97  |     |    |      |    | 2  | 1  |    |    |    | 1  | 134   |
| 3b      |       | 26 | 59  |     | 1  |      | 1  | 1  |    |    |    |    | 1  | 89    |
| 4a      |       | 35 | 100 |     |    |      | 1  | 1  | 3  |    |    |    | 1  | 141   |
| 4b      |       | 31 | 105 | 4   |    |      | 1  | 1  | 2  |    |    |    | 1  | 145   |
| 5a      |       | 13 | 107 | 4   |    |      |    | 1  | 2  |    |    |    | 1  | 128   |
| 5b      |       | 9  | 116 | 1   |    |      |    | 1  | 2  |    |    |    |    | 129   |
| 6a      |       |    | 130 | 14  |    |      |    |    | 2  |    |    |    | 1  | 147   |
| 6b      |       |    | 127 | 5   |    |      |    |    | 3  |    |    |    |    | 135   |
| 7a      |       |    | 118 | 12  |    |      |    | 1  | 2  |    |    |    |    | 133   |
| 7b      |       |    | 107 | 7   |    | 1    |    |    | 1  |    |    |    |    | 116   |
| 9a      |       |    | 102 | 114 | 3  | 1    |    |    | 2  |    |    |    | 1  | 223   |
| 9b      |       |    | 72  | 16  | 1  |      |    |    | 3  |    |    |    |    | 92    |

Table 4 shows the number of learner sessions per month for the recorded lectures of the second part of the course. At first sight, the pattern seems different. With respect to the recorded lectures for the second part of the course, the learner sessions are spread out over both the first two months after the recording is made, but also in the two months before the exam resits in 2010-04. The recorded lectures of the first part of the course have much fewer learner sessions for the retests.

**Table 4** Unique learner sessions per recorded lecture per month for the second part of the course

| Lecture | Month |    |    |    |    |      |    |    |    |    |    |    |    | Total |
|---------|-------|----|----|----|----|------|----|----|----|----|----|----|----|-------|
|         | 2009  |    |    |    |    | 2010 |    |    |    |    |    |    |    |       |
|         | 08    | 09 | 10 | 11 | 12 | 01   | 02 | 03 | 04 | 05 | 06 | 07 | 08 |       |
| 8a      |       |    | 63 | 92 | 34 | 59   | 15 | 16 | 23 | 1  |    |    | 1  | 304   |
| 8b      |       |    | 33 | 73 | 37 | 56   | 13 | 13 | 23 | 1  | 2  |    | 1  | 252   |
| 10a     |       |    | 9  | 94 | 48 | 93   | 14 | 19 | 41 |    | 2  |    | 2  | 322   |
| 10b     |       |    | 5  | 69 | 43 | 90   | 11 | 18 | 30 |    | 1  |    | 1  | 268   |
| 11a     |       |    |    | 59 | 43 | 109  | 9  | 19 | 37 |    |    |    |    | 276   |
| 11b     |       |    |    | 41 | 44 | 112  | 7  | 21 | 40 |    |    |    |    | 265   |
| 12a     |       |    |    | 37 | 50 | 155  | 6  | 18 | 63 |    |    |    |    | 329   |
| 12b     |       |    |    | 26 | 35 | 114  | 5  | 14 | 44 |    |    |    |    | 238   |
| 13a     |       |    |    | 11 | 40 | 149  | 3  | 16 | 45 |    |    |    |    | 264   |
| 13b     |       |    |    | 11 | 33 | 134  | 2  | 12 | 55 |    | 1  |    | 1  | 249   |
| 14a     |       |    |    |    | 40 | 126  | 2  | 10 | 55 |    |    |    |    | 233   |
| 14b     |       |    |    |    | 29 | 109  | 2  | 8  | 47 |    |    |    |    | 195   |
| 15a     |       |    |    |    | 30 | 119  | 2  | 8  | 42 |    | 2  |    |    | 203   |
| 15b     |       |    |    |    | 26 | 119  | 1  | 7  | 43 |    |    |    |    | 196   |
| 16a     |       |    |    |    | 17 | 121  |    | 8  | 49 |    |    |    |    | 195   |
| 16b     |       |    |    |    | 16 | 108  |    | 5  | 47 |    |    |    |    | 176   |
| 17a     |       |    |    |    |    | 189  |    | 8  | 70 |    |    |    |    | 267   |
| 17b     |       |    |    |    |    | 164  |    | 7  | 50 |    |    |    |    | 221   |
| 18a     |       |    |    |    |    | 92   | 2  | 4  | 27 |    | 2  |    |    | 127   |

## 5.2 User analysis

During the period August 2009 through August 2010, a total of 280 students (221 male and 59 female) participated actively in the course, meaning that they had a registered result for C01P01, C01P02 or both; one hundred ninety-five students (155 male and 40 female) or 69.6% of the total group completed the C01 course during that period.

During our study period, the recorded lectures for this course were viewed in a total of 2,650 learner sessions by a total of 291 unique students. Thirty-nine students (13.5%) viewed the recorded lectures but did not have a registered result for either C01P01 or C01P02. Of those 39 students, 29 only had two or fewer learner sessions. Four students had a total of seven learner sessions and one student had nine learner sessions without a registered result for either part of the course.

A total of 28 students (10%) participated in the course but never watched a single recorded lecture. Of those students, 11 students (39.3%) completed the C01 course successfully during our study period.

### 5.3 *Learner sessions analysis*

During the learner sessions analysis, the 2,650 learner sessions by 291 unique students were analysed in more detail. We used three questions as the basis for this analysis:

- Do students use the recorded lectures as a structural substitute for lecture attendance?
- Do students repeatedly use the recorded lectures or just once?
- Do students use the recorded lectures to study for the tests/exams for this course?

### 5.4 *Recorded lectures as a substitute for lecture attendance*

There was no direct method available in the data recordings to analyse whether students used the recorded lectures as a substitute for lecture attendance for the C01 course. Lecture attendance was not mandatory and no attendance register was created during the C01 course for 2009-2010. Instead, we used the dataset to analyse the students' viewing behaviour. If students used recorded lectures as a substitute for lecture attendance, they will be more likely to watch the full length of a recorded lecture. We assume that they have watched the full length of a recorded lecture if they received at least 80% of the video for the recorded lecture.

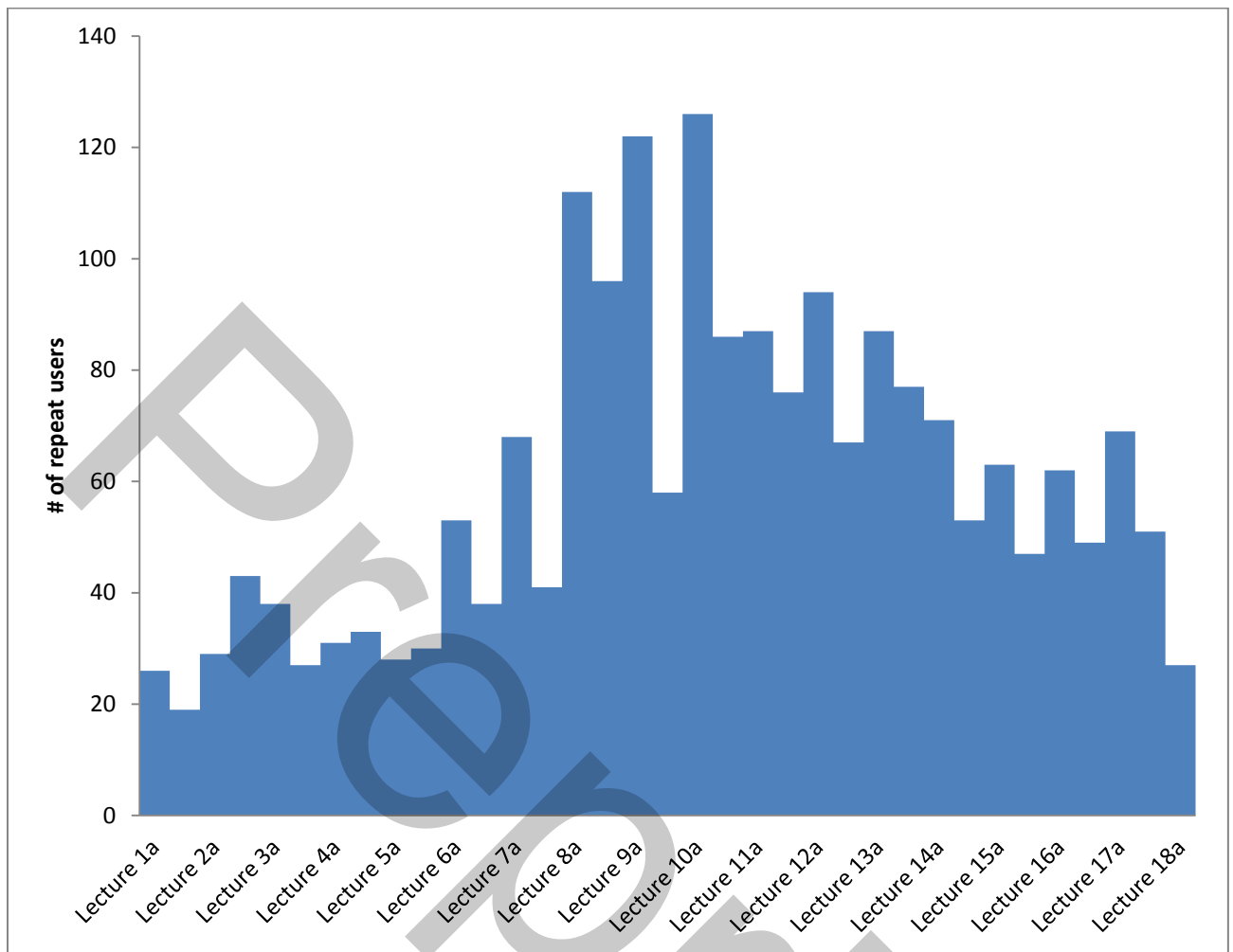
On average, each recorded lecture for the C01 course is watched completely 11 times during the one-year period covered by the dataset. Most of the students never watched the full length of a recorded lecture for this class. Of all students, only 27% watched the full length of one or more of the recorded lectures. The maximum number of recorded lectures watched in full length, for a single student for the C01 course, is 20 recorded lectures out of a total of 34 successfully recorded lectures. There were 13 students who watched 10 or more recorded lectures in full.

### 5.5 *Repeated use of lecture recordings*

Only 4% of the students watched all the recorded lectures during at least one learner session. On average, students watch almost half of all the recorded lectures during at least one learner session. The maximum number of times that a single recorded lecture had been viewed by a single student is 11 times.

Figure 7 shows the number of students per recorded lecture who viewed the recorded lecture during more than one learner session. Eighty-three percent watched a recorded lecture during more than one learner session, at least once. The recorded lectures during the second half of the course show a larger number of students that view the recordings more than once compared to the first half of the course.

**Figure 7** Repeated use per recorded lecture



### 5.6 Use of recorded lectures to study for the exam

In the survey reported by Gorissen, van Bruggen and Jochems (submitted), 58.2% of the students reported they usually watch a lecture recording while preparing for exams/tests. The graph in Figure 6 shows that there are four weeks in which there is an above-average use of the recorded lectures: during the week of the written test for C01P01 (on 26-10-2009), during the week that the assignment for C01P01 is due (10-11-2009), during the week leading up to the laptop test for C01P02 (23-1-2010), and during the week in which the retests for both the written test for C01P01 (12-4-2010) and the laptop test for C01P02 (16-4-2010) take place. This suggests that students do, in fact, use the recorded lectures to prepare for the tests.

We combined the data of the learner sessions during the two weeks before a test with the students that took the tests. For C0P01, we could only analyse the results for the written test. The results for the assignment for C01P01 are not registered as individual scores in the system.

**Table 5** Test results versus use of recorded lectures for the exam

| Date       | Test | Course part | Watched* |      | Did not watch* |      | <i>P</i> ** |
|------------|------|-------------|----------|------|----------------|------|-------------|
|            |      |             | Pass     | Fail | Pass           | Fail |             |
| 2009-10-26 | T1   | C01P01      | 132      | 9    | 80             | 22   | .0007       |
| 2010-01-23 | T2   | C01P02      | 90       | 75   | 6              | 51   | <.0001      |
| 2010-04-12 | T3   | C01P02      | 45       | 14   | 43             | 53   | .0001       |
| 2010-04-16 | T4   | C01P01      | 7        | 0    | 5              | 7    | .0174       |

\*1-14 days before the test

\*\*two-tailed Fisher's exact test

Table 5 shows the results for this analysis. Row T1 is the written test for C01P01 on 26-10-2009, row T2 is the laptop test for C01P02 on 23-1-2010, row T3 is the second opportunity to take the laptop test for C01P02, and finally, T4 is the second opportunity to take the written test for C01P01. Fifty-eight percent of all students have at least one learner session during the two weeks before a test.

We performed a two-tailed Fisher's exact test on the results for the four tests. For all four tests, the number of students that pass the test while having watched the recorded lectures is significantly higher than the expected number.

## 6 Conclusions

There are a number of steps required to clean and combine the available data into a dataset that can be used to analyse students' use of recorded lectures. In doing so, we were able to eliminate data that was not relevant for our research question. This also helped to obtain a clearer understanding of the students' actions while using the recorded lectures. By defining the users' sessions, the difference between the amount of received video and the learner session length showed that a learner session consists of more than just watching video. The data does not tell what exactly a student does during that time, however.

The analysis shows that it is useful to perform a triangulation of the survey data and the data logged by the system because there are differences between both data sets. The learner session analysis, for example, does not confirm the comment by students that they used the recorded lectures as a structural substitute for lecture attendance. Students do sometimes watch the full recorded lecture, but for the C01 course, none of students watched all recorded lectures in full and only a few of them watched all of the recorded lectures. The analysis does seem to confirm the remarks by students that they re-use recorded lectures from an earlier year if the course isn't re-recorded or if they prefer the recordings due to the lecturer that was recorded. Twenty-three percent of all learner sessions are on recorded lectures that have been recorded in the previous academic year.

The usage analysis shows that the moment of use of the recorded lectures is influenced more by the moments at which students study for the exams than by the moments at which the lectures are scheduled. We found strong evidence that studying the recorded lectures during exam preparation

increases the chances of passing the exam. This does not, however, preclude the hypothesis that these students were generally more active and more involved in the course overall.

If we want to support student navigation of recorded lectures, we should start by improving the options they have to navigate during exam preparation. We also should take into account that students do not just navigate within a single recorded lecture, but that most of them also navigate between different recorded lectures.

## 7 Acknowledgements

This research has been funded by Fontys University of Applied Sciences. The authors would like to thank the Student Office of TU/e for their assistance in this analysis and access to the dataset.

## 8 References

- Abowd, G., Atkeson, C., Brotherton, J., Enqvist, T., Gulley, P. and Lemon, J. (1998a) 'Investigating the capture, integration and access problem of ubiquitous computing in an educational setting', *SIGCHI conference on Human factors in computing systems*, Los Angeles, California: ACM Press/Addison-Wesley Publishing Co.
- Abowd, G., Atkeson, C., Feinstein, A., Hmelo, C., Kooper, R., Long, S., Sawhney, N. and Tani, M. (1996) 'Teaching and learning as multimedia authoring: The classroom 2000 project', *Fourth ACM International conference on Multimedia*. Boston, Massachusetts: ACM.
- Abowd, G., Brotherton, J. and Bhalodia, J. (1998b) 'Classroom 2000: A system for capturing and accessing multimedia classroom experiences', *CHI 98 conference on human factors in computing systems*. Los Angeles, California: ACM.
- Advanced Distributed Learning. (2004) SCORM Run-Time Environment Version 1.3.
- Baker, R. S. J. D., Penelope, P., Eva, B. and Barry, M. (2010) 'Data mining', *International Encyclopedia of Education*. Oxford: Elsevier.
- Behr, A. L. (1988) 'Exploring the lecture method: An empirical study', *Studies in Higher Education*, Vol. 13, pp. 189-200.
- Brotherton, J. and Abowd, G. (2004) 'Lessons learned from eClass: Assessing automated capture and access in the classroom', *ACM Transactions on Computer-Human Interaction (TOCHI)*, Vol. 11, pp. 121 – 155.
- Burby, J., Brown, A. and Committee, W. S. (2007) *Web Analytics Definitions*, Washington, DC: Web Analytics Association.
- Denzin, N. (2006) *Sociological Methods: A Sourcebook*, Piscataway, New Jersey: Transaction Publishers.
- Gorissen, P., Van Bruggen, J. M. and Jochems, W. (Submitted) 'Students and recorded lectures: Survey on current use and demands for higher education'.
- Gosper, M., Green, D., Mcneill, M., Phillips, R., Preston, G. and Woo, K. (2008) 'The impact of web-based lecture technologies on current and future practices in learning and teaching', Strawberry Hills, NSW: Australian Learning and Teaching Council.
- Koyun, H. (2007) Logging model for Windows Media Services. Microsoft Corporation. Available: <http://www.microsoft.com/windows/windowsmedia/howto/articles/loggingmodel.aspx> [Accessed 01-09-2009].
- Leoni, K. and Lichti, S. (2009) *Lecture Capture in Higher Education*, Evanston, IL: Northwestern University.
- Sheard, J. (2011) 'Basics of statistical analysis of interactions. Data from web-based learning environments', In: Romero, C., Ventura, S., Pechenizkiy, M. and Baker, R. S. J. D. (eds.) *Handbook of Educational Data Mining*, Boca Raton, FL: Chapman and Hall/CRC Press.
- Sonicfoundry. (n.d.) MediaSite. Available: <http://www.sonicfoundry.com/mediasite/> [Accessed 11/15/2010].



Traphagan, T. (2005) *Class Lecture Webcasting, Fall 2004 and Spring 2005: A Case Study*, Austin: University of Texas.

Traphagan, T. (2006) *Class Lecture Webcasting, Fall 2004, Spring 2005, and Fall 2005: Summary of Three Case Studies*, Austin: University of Texas.

Veeramani, R. and Bradly, S. (2008) *Insights Regarding Undergraduate Preference for Lecture Capture*, Madison, WI: University of Wisconsin-Madison E-Business Institute.

Witten, I. H. and Frank, E. 2002. 'Data mining: Practical machine learning tools and techniques with Java implementations', *ACM SIGMOD Record*, Vol. 31, pp. 76-77.

Preprint